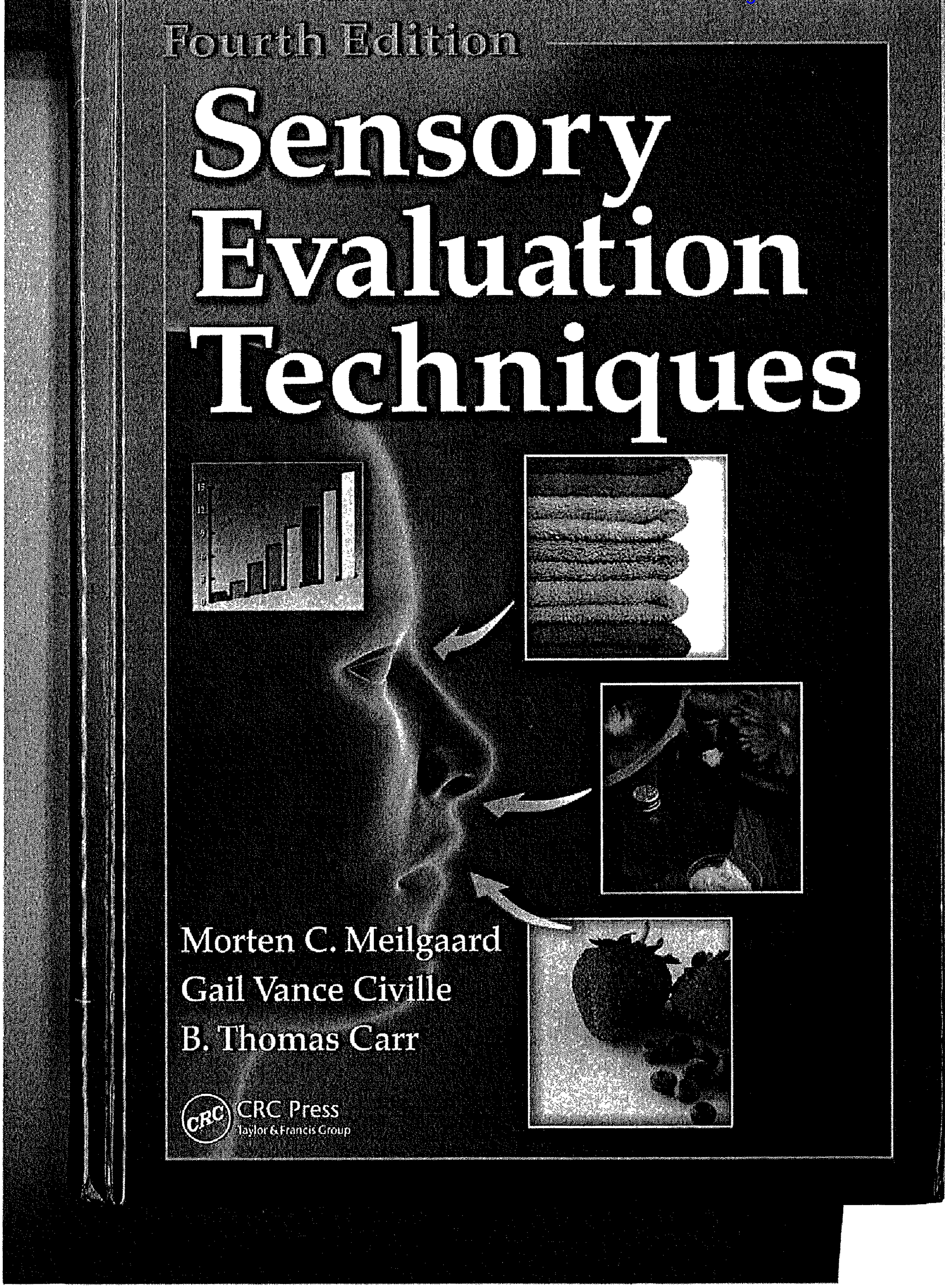
EXHIBIT 2


Fourth Edition
Sensory Evaluation Techniques
Morten C. Meilgaard
Gail Vance Civille
B. Thomas Carr
CRC Press
Taylor & Francis Group

The organization of the chapters and sections is also straightforward. Chapter 1 lists the steps involved in a sensory evaluation project, and Chapter 2 briefly reviews the workings of our senses. In Chapter 3, we list what is required of the equipment, the tasters, and the samples; while in Chapter 4, we have collected a list of those psychological pitfalls that invalidate many otherwise good studies. Chapter 5 discusses how sensory responses can be measured in quantitative terms. In Chapter 6, we describe all the common sensory tests for difference, the Triangle, Duo–trio, etc.; and, in Chapter 7, the various attribute tests, such as ranking and numerical intensity scaling, are discussed. Thresholds and just-noticeable differences are briefly discussed in Chapter 8, followed by what we consider the main chapters: Chapter 9 on selection and training of tasters, Chapters 10 and 11 on descriptive testing, and Chapter 12 on affective tests (consumer tests). All the descriptive references have been reviewed and revised for the Spectrum references in Chapter 11. Chapter 12 defines, in detail, several qualitative and quantitative classic methods for testing with consumers and includes substantial reviews of "fuzzy front end" and internet research techniques.

The body of text on statistical procedures is found in Chapters 13 and 14, but, in addition, each method (Triangle, Duo–trio, etc.) in Chapters 6 and 7 is followed by a number of examples showing how statistics are used in the interpretation of each. Basic concepts for tabular and graphical summaries, hypothesis testing, and the design of sensory panels are presented in Chapter 13. We refrain from detailed discussion of statistical theory, preferring instead to give examples. Chapter 14 discusses multifactor experiments that can be used, for example, to screen for variables that have large effects on a product, to identify variables that interact with each other in how they affect product characteristics, or to identify the combination of variables that maximize some desirable product characteristic such as consumer acceptability. Chapter 14 also contains a discussion of multivariate techniques that can be used to summarize large numbers of responses with fewer, meaningful ones to identify relationships among responses that might otherwise go unnoticed, and to group respondents of samples that exhibit similar patterns of behavior. New in the fourth edition is an overview of Thurstonian Scaling. In addition to studying differences among products, Thurstonian Scaling can be used to uncover the decision processes used by assessors during their evaluations of products. Also new in the fourth edition is a detailed discussion of data-relationship techniques used to link data from diverse sources collected on the same set of samples. The techniques are used to identify relationships, for example, between instrumental and sensory data or between sensory and consumer data. They can reveal the sensory and instrumental characteristics of products that have the greatest impact on acceptance and the intensities of these characteristics that are predicted to be most well liked by consumers.

At the end of the book, the reader will find guidelines for the choice of techniques and for reporting results plus the usual glossaries, indexes, and statistical tables.

With regard to terminology, the terms *assessor, judge, panelist, respondent, subject,* and *taster* are used interchangeably as are *he, she,* and *(s)he* for the sensory analyst (the sensory professional, the panel leader) and for individual panel members.

Senses, Journal of Sensory Studies, and *Journal of Texture Studies;* in the proceedings of the Pangborn Symposia (triennial) and the International Sensometrics Group (biannual), both usually published as individual papers in the journal *Food Quality & Preference;* and the proceedings of the Weurman Symposia (triennial, but published in book form, e.g., Martens, Dalen, and Russwurm 1987; Bessière and Thomas 1990). Sensory papers presented to the Institute of Food Technologists are usually published in the IFT's *Journal of Food Science* or *Food Technology.*

The methods that have been developed serve economic interests. Sensory testing can establish the worth of a commodity or even its very acceptability. Sensory testing evaluates alternative courses to select the one that optimizes value for money. The principal uses of sensory techniques are in quality control, product development, and research. They find application not only in characterization and evaluation of foods and beverages, but also in other fields such as environmental odors, personal hygiene products, diagnosis of illnesses, testing of pure chemicals, etc. The primary function of sensory testing is to conduct valid and reliable tests that provide data on which sound decisions can be made.

1.3 Human Subjects as Instruments

Dependable sensory analysis is based on the skill of the sensory analyst in optimizing the four factors, which we all recognize because they are the ones that govern any measurement (Pfenninger 1979).

1. Definition of the problem: what is to be measured must be precisely defined; important as this is in "hard" science, it is much more so with senses and feelings.
2. Test design: not only must the design leave no room for subjectivity and take into account the known sources of bias, but it also must minimize the amount of testing required to produce the desired accuracy of results.
3. Instrumentation: the test subjects must be selected and trained to give a reproducible verdict; the analyst must work with them until he/she knows their sensitivity and bias in the given situation.
4. Interpretation of results: using statistics, the analyst chooses the correct null hypothesis and the correct alternative hypothesis, and draws only those conclusions that are warranted by the results.

Tasters, as measuring instruments, are (1) quite variable over time; (2) very variable among themselves; and (3) highly prone to bias. To account adequately for these shortcomings requires (1) that measurements be repeated, (2) that enough subjects (often 20–50) are made available so that verdicts are representative, and (3) that the sensory analyst respects the many rules and pitfalls that govern panel attitudes (see Chapter 4). Subjects vary innately in sensitivity by a factor of 2–10 or more (Meilgaard and Reid 1979; Pangborn 1981) and should not be interchanged halfway through a project. Subjects must be selected for sensitivity and must be trained and retrained (see Chapter 9) until they fully understand the task at hand. The annals of sensory testing are replete with results that are unreliable because many of the panelists did not understand the questions and/or the terminology used in the test, did not recognize the flavor or texture parameters in the products, or did not feel comfortable with the mechanics of the test or the numerical expressions used.

2. *Determine the test objective.* Once the objective of the project can be clearly stated, the sensory analyst and the project leader can determine the test objective. overall difference, attribute difference, relative preference, acceptability, etc. Avoid attempting to answer too many questions in a single test. A good idea is for the sensory analyst and project leader to record, in writing, before the test is initiated the project objective, the test objective, and a brief statement of how the test results will be used.
3. *Screen the samples.* During the discussion of project and test objectives, the sensory analyst should examine all of the sensory properties of the samples to be tested. This enables the sensory analyst to use test methods that take into account any sensory biases introduced by the samples. For example, visual cues (color, thickness, sheen) may influence overall difference responses, such as those provided in a triangle test, e.g., to measure differences due to sweetness of sucrose vs. aspartame. In such a case, an attribute test would be more appropriate. In addition, product screening provides information on possible terms to be included in the scoresheet.
4. *Design the test.* After defining the project and test objectives and screening the samples, the sensory analyst can proceed to design the test. This involves selection of the test technique (see Chapter 6 through Chapter 8, Chapter 10 through Chapter 12, and Chapter 15); selecting and training subjects (see Chapter 9); designing the accompanying scoresheet (ballot, questionnaire); specifying the criteria for sample preparation and presentation (see Chapter 3); and determining how the data will be analyzed (see Chapter 13 and Chapter 14). Care must be taken, in each step, to adhere to the principles of statistical design of experiments to ensure that the most sensitive evaluation of the test objective is attained.
5. *Conduct the test.* Even when technicians are used to carry out the test, the sensory analyst is responsible for ensuring that all the requirements of the test design are met.
6. *Analyze the data.* Because the procedure for analysis of the data was determined at the test design stage, the necessary expertise and statistical programs, if used, will be ready to begin data analysis as soon as the study is completed. The data should be analyzed for the main treatment effect (test objective) as well as other test variables, such as order of presentation, time of day, different days, and/or subject variables such as age, sex, geographic area, etc.
7. *Interpret and report results.* The initial clear statement of the project and test objectives will enable the sensory analyst to review the results, express them in terms of the stated objectives, and make any recommendations for action that may be warranted. The latter should be stated clearly and concisely in a written report that also summarizes the data, identifies the samples, and states the number and qualification of subjects (see Chapter 16).

The main purpose of this book is to help the sensory analyst develop the methodology, subject pool, facilities, and test controls required to conduct analytical sensory tests with trained and/or experienced tasters. In addition, Chapter 12 discusses the organization of consumer tests, i.e., the use of naïve consumers (nonanalytical) for large-scale evaluation, structured to represent the consumption and responses of a large population of the product market.

The role of sensory evaluation is to provide valid and reliable information to R&D, production, and marketing in order for management to make sound business decisions about the perceived sensory properties of products. The ultimate goal of any sensory

has reached the higher level of training, it can tackle many types of samples without the need for separate training sessions for each.

5.2 Psychophysical Theory

Psychophysics is a branch of experimental psychology devoted to studying the relationships between sensory stimuli and human responses, i.e., to improving understanding of how the human sensory system works. University psychophysicists are constantly refining the methods by which a response can be measured, and sensory analysts need to study their techniques and cooperate in their experiments. This chapter will provide an overview of psychophysics as applied to sensory testing. Those interested in more detail should read the references listed at the beginning of Section 5.1; see also Lawless (1990).

A major focus of psychophysics is to discover the form of the psychophysical function, the relationship between a stimulus, C, and the resulting sensation, R, preferably expressed as a mathematical function, $R=f(C)$ (see Figure 5.2).

While the stimulus is either known (an added concentration) or easy to measure (a peak height, an Instron reading), it is the sensation that causes difficulty. The subject must be asked questions and given instructions such as:

> Judge this odor on a scale of 0–99,
>
> Is this sensation 2× as strong or 3× as strong?
>
> Which of these solutions has the strongest taste of quinine?

No one, however, can answer such questions reproducibly and precisely. A variety of experimental techniques are being used, e.g., comparison with a second, better known sensation such as the loudness of a tone (this is called cross-modality matching, see p. 59),

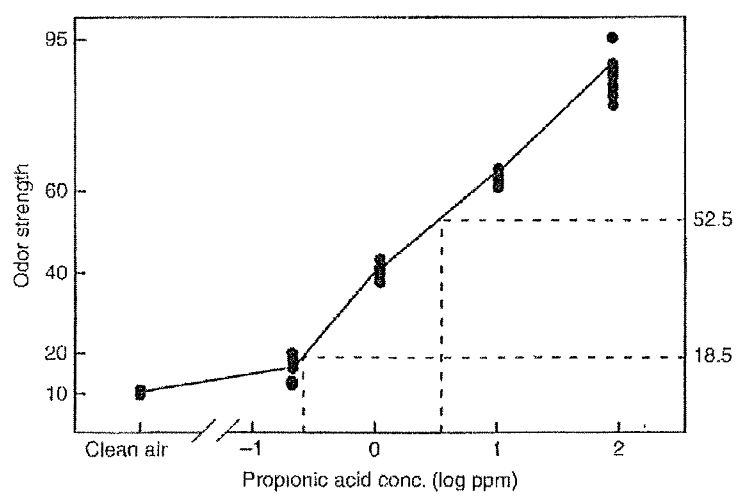


FIGURE 5.2
Example of a psychophysical function. Odor strength was rated 0–99 with zero = no odor or nasal irritation sensation. (From Kendal-Reed et al. 1998. *Chemical Senses*, Vol. 23, Oxford University Press, 71–82. With permission.)

Compared with difference testing, scaling is a more informative—and therefore a more useful—form of recording the intensity of perception. As with ranking, the results are critically dependent on how well the panelists have been familiarized with the attribute under test and with the scale being used. In this respect, three different philosophies have been applied (Muñoz and Civille 1998):

- Universal scaling, in which panelists consider all products and intensities they have experienced as their highest intensity reference point (example: the Spectrum aromatics scale uses the cinnamon impact of Big Red chewing gum as a 13 in intensity on a 15-point scale).
- Product-specific scaling, in which panelists consider only their experience within the selected product category in setting their highest reference point (example: the vanilla impact of typical vanilla cookies was set at 10 on a 15-point product specific scale).
- Attribute-specific scaling, in which panelists consider their experience of the selected attribute across all products in setting their highest reference point (example: a specific toothpaste is assigned the top value of 13 for the peppermint aromatic in any product).

A common problem with scales is that panelists tend to use only the middle section. For example, if ciders are judged for intensity of "appley" flavor on a scale of 0–9, subjects will avoid the numbers 0, 1, and 2 because they tend to keep these in reserve for hypothetical samples of very low intensity, which may never come. Likewise, the numbers 7, 8, and 9 are avoided in anticipation of future samples of very high intensity, which may never come. The result is that the scale is distorted. For example, a cider of outstanding apple intensity may be rated 6.8 by the panel while a cider that is only just above the average may receive a 6.2.

Although the properties of data obtained from any response scale may vary with the circumstances of the test (e.g., experience of judges in the test, familiarity of the attribute), it is typically assumed that:

- Category scaling (ISO term: *rating*) yields ordinal or interval data;
- Line scaling (ISO term: *scoring*) yields interval data;
- Magnitude estimation scaling (often called *ratio scaling*) sometimes, but not always, yields ratio data.

5.6.1 Category Scaling

A category (or partition) scale is a method of measurement in which the subject is asked to "rate" the intensity of a particular stimulus by assigning it a value (category) on a limited, usually numerical, scale. Category scale data are generally considered to be at least ordinal-level data. They do not generally provide values that measure the degree (how much) one sample is more than another. On a 7-point category scale for hardness, a product rated a 6 is not necessarily twice as hard as a product with a 3 hardness rating. The hardness difference between 3 and 6 may not be the same as that between 6 and 9. Although attempts are made to encourage panelists to use all intervals as equal, panelists may also tend to use the categories with equal frequency, except that they usually avoid the use of the two scale endpoints to

respect to overall differences and attribute differences (difference panels) or full descriptions of product standards, product changes over time, or ingredient and processing manipulation, and for construction and interpretation of consumer questionnaires (descriptive panels). The sensory analyst must also define the resources required to develop and maintain a sensory panel system.

9.2.1 Personnel

Heading the list of resources required is (1) a large enough pool of available candidates from which the panel can be selected (see Appendix 9.2A for possible sources for panel candidates), (2) a sensory staff to implement the selection, training, and maintenance procedures, including a panel leader and technician, and (3) a qualified person to conduct the training process. Ideally, panelists should come from within the organization, as they are located at the site where the samples are prepared (e.g., R&D facility or plant). Before a descriptive panel is trained, consideration is given to the choice of a panel leader. An effective panel leader is a person who is able to serve as the connection between product developers or other panel clients. The panel leader works with the panel to ensure that the panel has a clear understanding of attributes and scales as well as the ability to translate the panel data into actionable information. A successful panel leader is also a person who (1) has knowledge of sensory attributes; (2) has good group dynamic skills; (3) has listening and or attending skills; (4) is creatively alert; and (5) is patient. A panel leader may come from the panel itself. If this is the case, the panel leader should be additionally trained to manage the panel and communicate with the research team so that the information provided to product developers and other scientists is reliable, valid, and useful. If a panel is large, a panel technician may also be required to be responsible for all sample procurement, preparation, and presentation, as well as completing all the necessary documentation of the panel protocol and data output.

Some companies choose to test products at a different site, which may be another company facility. With reduced laboratory staffing, many companies have opted to use residents recruited from the local community as panelists rather than bench chemists and support staff from the labs. Outside panelists may be available for more hours per week, and may be cheaper and more focused for longer panel sessions. The primary drawbacks of external panelists are that they often require more time and effort to train in the technical aspects of panel work, and that they do not provide the inherent proprietary security of internal employees.

Panel candidates and management must understand, in advance, the amount of time required (personnel hours) for the selection and training of the particular panel in question. An assessment of the number of hours needed for panelists, technicians, and a panel leader should be presented and accepted before the development process is initiated. The individual designated to select and train the panel is often a member of the sensory staff who is experienced and trained in the specific selection and training techniques needed for the challenge at hand.

9.2.2 Facilities

The physical area for the selection, training, and ongoing work of a panel must be defined before development of the panel begins. A training room and panel testing facilities (booths and/or round table, conference room, etc.) must have the proper environmental controls (see Chapter 3), be of sufficient size to handle all of the panelists and products projected, and be located near the product preparation area and panelist pool.

TABLE 9.4

Suggested Materials for Ranking/Rating Tests

Taste	Sensory Stimuli	Concentration			
Sour	Citric acid/water, g/L	0.25	0.5	1.0	1.5
Sweet	Sucrose/water, g/L	10	20	50	100
Bitter	Caffeine/water, g/L	0.3	0.6	1.3	2.6
Salty	Sodium chloride/water, g/L	1.0	2.0	5.0	10
Odor					
Alcoholic	3-Methylbutanol/water, mg/L	10	30	80	180
Texture					
Hardness	Cream cheese,[a] American cheese,[a] peanuts, carrot slices[a]				
Fracturability	Corn muffin,[a] graham cracker, Finn crisp bread, life saver				

[a] At 1/4 inch thickness.

samples in ascending order (or rate them using the prescribed scale) according to the level of the stated attribute (sweetness, oiliness, stiffness, surface smoothness, etc.); see suggested materials in Table 9.4.

Typical scoresheets are shown in Table 9.5 and Table 9.6 The selection sequence may make use of more than one attribute ranking/rating test, especially if the ultimate panel will need to cover several sense modalities, e.g., color, visual surface oiliness, stiffness, and surface smoothness.

9.3.1.4 *Interpretation of Results of Screening Tests*

Matching tests. Reject candidates scoring less than 75% correct matches. Reject candidates for attribute tests who score less than 60% in choosing the correct descriptor.

Detection/discrimination tests. When using triangle tests, reject candidates scoring less than 60% on the "easy" tests (6 × threshold) or less than 40% on the "moderately difficult" tests (3 × threshold). When using duo-trio tests, reject candidates scoring less than 75% on the easy tests or less than 60% on the moderately difficult tests. Alternatively, use the sequential tests procedure, as described in Chapter 6, p. 100.

Ranking/rating tests. Accept candidates ranking samples correctly or inverting only adjacent pairs. In the case of rating, use the same rank-order criteria and expect candidates to use a large portion of the prescribed scale when the stimulus covers a wide range of intensity.

9.3.2 Training

To ensure development of a professional attitude to sensory analysis on the part of panelists, conduct the training in a controlled professional sensory facility. Instruct subjects

TABLE 9.5

Scoresheet, Ranking Test for Intensity

	Code
Least salty	______
	______
	______
Most salty	______

Rank the salty taste solutions in the coded cups in ascending order of saltiness.

on	
.0	1.5
)	100
l.3	2.6
5.0	10
)	180

ording to the level
othness, etc.); see

ion sequence may
the ultimate panel
ness, stiffness, and

. Reject candidates
criptor.
idates scoring less
oderately difficult"
ng less than 75% on
ternatively, use the

inverting only adja-
ect candidates to use
e range of intensity.

on the part of pane-
ty. Instruct subjects

TABLE 9.6

Scoresheet, Rating Test for Intensity

Code	
463	None________________Strong
318	None________________Strong
941	None________________Strong
502	None________________Strong

Rate the saltiness of each coded solution for intensity/strength of saltiness using the line scale for each

how to precondition the sensory modality in question, e.g., not to use perfumed cosmetics and to avoid exposure to foods or fragrances for 30 min before sessions; how to prepare skin or hands for fabric and skinfeel evaluations; and how to notify the panel leader of allergic reactions that affect the test modality. On any day, excuse subjects suffering from colds, headaches, lack of sleep, etc.

From the outset, teach subjects the correct procedures for handling the samples before and during evaluation. Stress the importance of adhering to the prescribed test procedures, reading all instructions, and following them scrupulously. Demonstrate ways to eliminate or reduce sensory adaptation, e.g., taking shallow sniffs of fragrances and leaving several tens of seconds between sample evaluations. Stress the importance of disregarding personal preferences and concentrating on the detection of difference.

Begin by presenting samples of the product(s) under study that represent large, easily perceived sensory differences. Concentrate initially on helping panelists to understand the scope of the project and to gain confidence. Repeat the test method using somewhat smaller but still easily perceived sample differences. Allow the panel to learn through repetition until full confidence is achieved.

For attribute difference tests, carefully introduce panelists to the attributes, the terminology used to describe them, and the scale method used to indicate intensity. Present a range of products showing representative intensity differences for each attribute.

Continue to train "on the job" by using the new panelists in regular discrimination tests. Occasionally, introduce training samples to simulate "off-notes" or other key product differences to keep the panel on track and attentive.

Be aware of changes in attitude or behavior on the part of one or more panelists who may be confused, losing interest, or distracted by other problems. The history of sensory testing is full of incredible results that could have come only from panelists who were "lost" during the test with the sensory analyst failing to anticipate and detect a failure in the "test instrument."

9.4 Selection and Training of Panelists for Descriptive Testing

9.4.1 Recruiting Descriptive Panelists

Panelist recruiting is a key element in creating a successful descriptive panel (Appendix 9.2A). A descriptive panel describes products using attributes and intensities, so panelists must be capable of using both terms and expressions of magnitude to "tell the story" of the products. Even though a descriptive panelist should be a discriminator, it is important that the panelist also has proven abilities to think and communicate.

Step one in panel building is recruiting as many interested, potential panelists as possible They must be informed of some of the details surrounding the study and what the benefits are for them (money, knowledge, expertise, etc.) Postings, newspapers ads, and announcements on radio or during public events name only a few pathways. The postings and newspaper ads should be placed where they are most likely to be seen by people who are interested in food, beauty, and home (Stoer, Rodriguez, and Civille 2002). Figure 9.1 is an example of such an ad.

9.4.2 Selection for Descriptive Testing

When selecting panelists for descriptive analysis, the panel leader or panel trainer should determine each candidate's capabilities in three major areas:

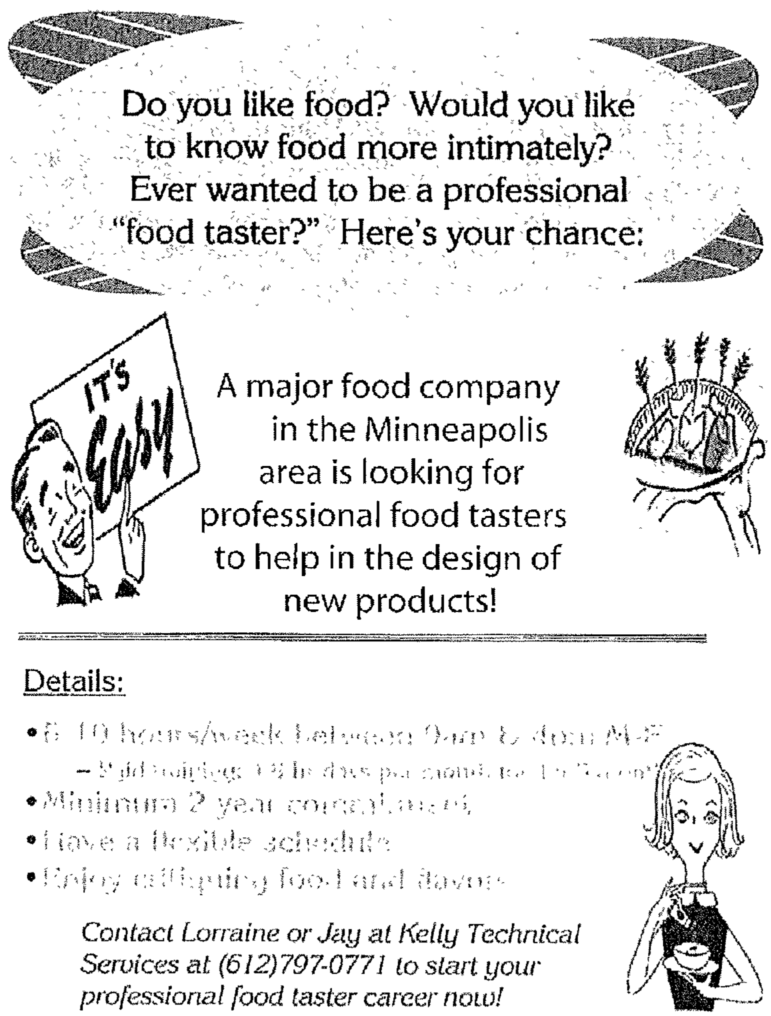


FIGURE 9.1
Example of a descriptive panel recruiting advertisement (Stoer, Rodriguez, and Civille 2002).

1. For each of the sensory properties under investigation (such as fragrance odor, flavor, oral texture, handfeel or skinfeel), the ability to detect differences in characteristics present and in their intensities.
2. The ability to describe those characteristics using (a) verbal descriptors for the characteristics and (b) scaling methods for the different levels of intensity.
3. The capacity for abstract reasoning, as descriptive analysis depends heavily upon the use of references when characteristics must be quickly recalled and applied to other products.

In addition to screening panelists for these descriptive capabilities, panel leaders must prescreen candidates for the following personal criteria:

1. Interest in full participation in the rigors of the training, practice, and ongoing work phases of a descriptive panel.
2. Availability to participate in 80% or more of all phases of the panel's work; whether conflict with home life, work load, travel, or even the candidate's supervisor may eventually cause the panelist to drop off the panel during or after training, thus losing one panelist from an already small number of 10–15.
3. General good health and no illnesses related to the sensory properties being measured, such as:
   a. Diabetes, hypoglycemia, hypertension, dentures, chronic colds or sinusitis, or food allergies in those candidates for flavor and/or texture analysis of foods, beverages, pharmaceuticals, or other products for internal use.
   b. Chronic colds or sinusitis, for aroma analysis of foods, fragrances, beverages, personal care products, pharmaceuticals, or household products.
   c. Central nervous system disorders or reduced nerve sensitivity due to the use of drugs affecting the central nervous system, for tactile analysis of personal care skin products, fabrics, or household products.

The ability to detect and describe differences, the ability to apply abstract concepts, and the degree of positive attitude and predilection for the tasks of descriptive analysis can all be determined through a series of tests that include:

- A set of prescreening questionnaires.
- A set of acuity tests.
- A set of ranking/rating tests.
- A personal interview.

The investment in a descriptive panel is large in terms of time and human resources, and it is wise to conduct an exhaustive screening process, rather than train unqualified subjects.

Lists of screening criteria for three descriptive methods (the Flavor Profile, Quantitative Descriptive Analysis, and Texture Profile) can be found in ASTM Special Technical Publication 758 (1981). The following criteria listed are those used to select subjects for training in the spectrum Method of descriptive analysis, as described in Chapter 11. These can be applied to the screening of employees, or for external screening in cases where recruiting from the local community is preferred due to the amount of time necessary (20–50 h per person per week). The additional prescreening questionnaires are used to select individuals who can verbalize and think conceptually. This reduces the risk of selecting outside

panelists who have sensory acuity but cannot acquire the "technical" orientation of panels recruited from inside the company.

9.4.2.1 Prescreening Questionnaires

For a panel of 15, typically 40–50 candidates may be prescreened using questionnaires such as those shown in Appendix 9.1. Appendix 9.1A applies to a tactile panel (skinfeel or fabric feel); Appendix 9.1B to a flavor panel; Appendix 9.1C to an oral texture panel; and Appendix 9.1D to a fragrance panel. Appendix 9.1E evaluates the candidate's potential to learn scaling and can be used with any of the preceding questionnaires in Appendix 9.1. Of the 40–50 original candidates, generally 20–30 qualify and proceed to the acuity tests.

9.4.2.2 Acuity Tests

To qualify for this stage, candidates should:

- Indicate no medical or pharmaceutical causes of limited perception.
- Be available for the training sessions.
- Answer 80% of the verbal questions in the prescreening questionnaires in Appendix 9.1A through Appendix 9.1D correctly and clearly.
- In the questionnaire Appendix 9.1E, assign scalar ratings that are within 10–20% of the correct value for all figures.

Candidates should demonstrate ability to:

- Detect and describe characteristics present in a qualitative sense.
- Detect and describe intensity differences in a quantitative sense.

Therefore, although detection tests (e.g., triangle or duo-trio tests using variations in formulation or processing of the product to be evaluated) may yield a group of subjects who can detect small product variables, detection alone is not enough for a descriptive panelist. To qualify, subjects must be able to adequately discriminate and describe some key sensory attributes within the modalities used within the product class under test, and also must show ability to use a rating scale correctly to describe differences in intensity.

Detection. The panel trainer presents a series of samples representing key variables within the product class, in the form of triangle or duo-trio tests (Zook and Wesmann 1977). Differences in process time or temperature (roast, bake, etc.), ingredient level (50% or 150% of normal), or packaging can be used as sample pairs to determine acuity in detection. Attempt to present the easier pairs of samples first and follow with pairs of increasing difficulty. Select subjects who achieve 50–60% correct replies in triangle tests, or 70–80% in duo-trio tests, depending on the degree of difficulty of each test.

Description. Present a series of products showing distinct attribute characteristics (fragrance/flavor oils, geometrical texture properties [Civille and Szczesniak 1973]) and ask candidates to describe the sensory impression. Use the fragrance list in Table 9.1 without a list of descriptors from which to choose. The candidate must describe each fragrance using his/her own words. These may include chemical terms (e.g., cinnamic aldehyde), common flavor terms (e.g., cinnamon), or related terms (e.g., like Red Hots candy, Big Red gum, and Dentyne). Candidates should be able to describe 80% of the

stimuli using chemical, common, or related terms and should at least attempt to describe the remainder with less specific terms (e.g., sweet, brown spice, hot spice).

9.4.2.3 *Ranking/Rating Screening Tests for Descriptive Analysis*

Having passed the prescreening tests and acuity tests, the candidate is ready for screening with the actual product class and/or sensory attribute for which the panel is being selected. A good example for a Camembert cheese panel is given by Issanchou, Lesschaeve, and Köster (1995). Candidates should rank or rate a number of products on a selection of key attributes, using the technique of the future panel. These tests can be supplemented with a series of samples that demonstrate increasing intensity of certain attributes, such as tastes and odors (see Table 9.4), or oral texture properties (Appendix 11.2, Texture Section D, Scale 5 is suitable, containing hardness standards from cream cheese = 1.0 to hard candy = 14.5; also Scale 10 that contains standards for crispness from Granola Bar at 2.0 to cornflakes at 14.0). A questionnaire such as Table 9.7 is suitable. For certain skinfeel and fabric feel properties, use Appendix 11.2E or Appendix 11.2F, or reference samples may need to be selected from among commercial products and laboratory prototypes that represent increasing intensity levels of selected attributes. Choose candidates who can rate all samples in the correct order for 80% of the attributes scaled. Allow for reversal of adjacent samples only, and check that candidates use most of the scale for at least 50% of the attributes tested.

9.4.2.4 *Personal Interview*

Especially for descriptive panels, a personal interview is necessary to determine whether candidates are well suited to the group dynamics and analytical approach. Generally, candidates who have passed the prescreening questionnaire and all of the acuity tests are interviewed individually by the panel trainer or panel leader. The objective of the interview is to confirm the candidate's interest in the training and work phases of the panel, including his/her availability with respect to work load, supervisor, and travel, and also communication skills and general personality. Candidates who express little interest in the sensory programs as a whole, and in the descriptive panel in particular, should be excused. Individuals with very hostile or very timid personalities may also be excluded, as they may detract from the needed positive input of each panelist.

TABLE 9.7

Scoresheet Containing Two Ranking Tests Used to Screen Candidates for a Texture Panel

Descriptive Texture Panel Screening

1. Place one piece of each product between molars; bite through once; evaluate for hardness. Rank the samples from least hard to most hard

Least hard ______________

____

____

____

Most hard ______________

2. Place one piece of each product between molars; bite down once and evaluate for crispness (crunchiness)

Least crisp ______________

____

____

____

Most crisp ______________

9.4.2.5 Mock Panel

Some companies further screen panelist candidates by inviting them to a "mock panel" at which they are asked to evaluate and comment on two or more products. Candidates are presented with the products, write down their perceptions (sensory parameters described by the session panel leader, e.g., "The flavor and texture of these crackers"). The panel leader then directs a discussion of the results that provides each panelist a time to express his or her perceptions. Observation of the panelists' behavior is helpful in deciding which candidates work best in a group, express concepts clearly, and participate in discussions of different perceptions.

9.4.3 Training for Descriptive Testing

The important aspect of any training sequence is to provide a structured framework for learning based on demonstrated facts and to allow the students, in this case panelists, to grow in both skills and confidence. Most descriptive panel training programs require between 40 and 120 h of training. The amount of time needed depends on the complexity of the product (wine, beer, and coffee panels require far more time than those evaluating lotions, creams or breakfast cereals), on the number of attributes to be covered (a short-version descriptive technique for quality control or storage studies, Chapter 11, p. 193, requires fewer and simpler attributes), and on the requirements for validity and reliability (a more experienced panel will provide greater detail with greater reproducibility).

9.4.3.1 Terminology Development

The panel leader or panel trainer, in conjunction with the project team, must identify key product variables to be demonstrated to the panel during the initial stages of training. The project team should prepare a prototype or collect an array of products from commercially available samples as a frame of reference that represents as many of the attribute differences likely to be encountered in the product category as possible. The panel is first introduced to the chemical (olfaction, taste, chemical feeling factors) and physical principles (rheological, geometrical, etc.) that govern or influence the perception of each product attribute. With these concepts and terms as a foundation, the panel then develops procedures for evaluation and terminology with definitions and references for the product class.

Examples of this process are discussed by Szczesniak and Kleyn (1963) for oral texture, Schwartz (1975) and Civille and Dus (1991) for skincare products, McDaniel et al. (1987) for wines, Meilgaard and Muller (1987) for beer, Lyon (1987) for chicken, Johnsen et al. (1988) for peanuts, Johnsen and Civille (1986) for beef, and Johnsen, Civille, and Vercellotti (1987) for catfish. Typically, the first stage of training may require 15–20 h as panelists begin to develop an understanding of the broad array of descriptors that fall into the category being studied (appearance, flavor, oral texture, etc.). This first phase is designed to provide them with a firm background in the underlying modality and for them to begin to perceive the different characteristics as they are manifest in different product types.

9.4.3.2 Introduction to Descriptive Scaling

The scaling method of choice may be introduced during the first 10–20 h of training. By using a set of products or references that represent three to five different levels of each attribute, the panel leader reinforces both the sensory characteristic and the scaling method by demonstrating different levels or intensities across several attributes. Appendix 11.2 provides examples of different intensity levels of several sensory attributes

to a "mock panel" at
ucts. Candidates are
arameters described
·ackers"). The panel
list a time to express
al in deciding which
ate in discussions of

ured framework for
his case panelists, to
g programs require
ds on the complexity
han those evaluating
be covered (a short-
Chapter 11, p. 193,
lidity and reliability
eproducibility).

n, must identify key
ages of training. The
s from commercially
ny of the attribute
ssible. The panel is
actors) and physical
e the perception of
tion, the panel then
s and references for

963) for oral texture,
IcDaniel et al. (1987)
icken, Johnsen et al.
iville, and Vercellotti
15–20 h as panelists
ors that fall into the
st phase is designed
nd for them to begin
ent product types.

-20 h of training. By
ferent levels of each
stic and the scaling
several attributes.
al sensory attributes

for several sensory descriptive categories: Flavor (aromatics, tastes, feeling factors), solid and semisolid texture (Muñoz 1986) (hardness, adhesiveness, springiness, etc.), skinfeel (ASTM 1997; Civille and Dus 1991) (wetness, slipperiness, oiliness, etc.), and fabric feel (Civille and Dus 1990) (slipperiness, grittiness, fuzziness, etc.).

The continued use of intensity reference scales during practice is meant to provide ongoing reinforcement of both attributes and intensities so that the panel begins to see the descriptive process as a use of terms and numbers (characteristics and intensities) to define or document any product in the category learned.

9.4.3.3 *Initial Practice*

The development of a precise lexicon for a given product category is often a three-step process. In the first step, a full array of products, prototypes, or examples of product characteristics are presented to the panel as a frame of reference. From this frame of reference, the panel generates an original long list of descriptors to which all panelists are invited to contribute. In the second stage, the original list, containing many overlapping terms, is rearranged and reduced into a working list in which the descriptors are comprehensive (they describe the product category completely) and yet discrete (overlapping is minimized). The third and last stage consists of choosing products, prototypes, and external references that can serve to represent good examples of the selected terms.

After the panel has a grasp on the terminology and a general understanding of the use of each scale, the panel trainer or leader presents a series of samples to be evaluated, one at a time, two or more of which represent a very wide spread in qualitative (attributes) and quantitative (intensity) differences. At this early stage of development, which lasts 15–40 h, the panel gains basic skills and confidence. The disparate samples allow the panel to see that the terms and scales are effective as descriptors and discriminators and help the members to gain confidence both as individuals and as a group.

9.4.3.4 *Small Product Differences*

With the help of the project/product team, the panel leader collects samples that represent smaller differences within the product class, including variations in production variables and/or bench modifications of the product. The panel is encouraged to refine the procedures for evaluation and the terminology with definitions and references to meet the needs of detecting and describing product differences. Care must be taken to reduce variations between supposedly identical samples; panelists in training tend to see variability in results as a reflection of their own lack of skill. Sample consistency contributes to panel confidence. This stage represents 10–15 h of panel time.

9.4.3.5 *Final Practice*

The panel should continue to test and describe several products during the final practice stage of training (15–40 h). The earlier samples should be fairly different, and the final products tested should approach the real-world testing situations for which the panel will be used.

During all five stages of the training program, panelists should meet after each session and discuss results, resolve problems or controversies, and ask for additional qualitative or quantitative references for review. This interaction is essential for developing the common terminology, procedures for evaluation, and scaling techniques that characterize a finely tuned sensory instrument.

9.5 Panel Performance and Motivation

Any good measuring tool needs to be checked regularly to determine its ability to perform validly and consistently. In the case of a sensory panel, the individuals, as well as the panel as a whole, need to be monitored. Panels are comprised of human subjects who have other jobs and responsibilities in addition to their participation in the sensory program; it is necessary to find ways to maintain the panelists' interest and motivation over long periods of product testing.

9.5.1 Performance

For both difference and descriptive panels, the sensory analyst needs to have a measure of the performance of each panelist and of the panel, in terms of validity and reproducibility. Validity is the correctness of the response. In certain difference tests, such as the triangle and duo-trio, and in some directional attribute tests, the analyst knows the correct answer (the odd sample, the coded reference, the sweeter sample) and can assess the number of correct responses over time. The percent of correct responses can be computed for each panelist on a regular monthly or bimonthly basis. Weighted scores can also be calculated, based on the difficulty of each test in which the panelist participated (Aust 1984). For the panel as a whole, validity can be measured by comparing panel results to other sensory test data, instrumental data, or the known variation in the stimulus, such as increased heat treatment, addition of a chemical, etc.

Reliability, or the ability to reproduce results, can be easily assessed for the individual panelists and for the panel as a whole by replicating the test, using duplicate test samples, or using blind controls.

For descriptive data that are analyzed statistically by the analysis of variance, the panelists' performance can be assessed across each attribute as part of the data analysis (see ASTM (1981), or Lea, Næs, and Rødbottenm (1997) for a detailed description of this analysis applied to a set of QDA results). It is recognized and accepted in QDA that panelists will use different parts of the scale to express their perceptions of the same sample. It is the relative differences in their ratings and not their absolute values that are considered important. In other descriptive methods, such as Spectrum, panelists are calibrated through the use of references to use the same part of the scale when evaluating the same sample. A descriptive panel of this type is equivalent to an analytical instrument that requires regular calibration checks. Several approaches, in addition to the ASTM guideline just mentioned, are appropriate for monitoring the individual and combined performance of "calibrated" panelists. Two aspects of performance that require monitoring are the panel's accuracy (bias) and its precision (variability). See also Nielsen, Hyldig, and Sørensen (2005).

Bias. To assess a panelist's ability to be "on target," the panel leader can determine the panelist's ability to match the accepted intensity of the attributes of a control or reference. The statistical measure of difference from the target or control rating, called *bias*, is defined as:

$$\text{panelist bias, } d = x - \mu, \qquad (9.1)$$

where d is the deviation or bias, x is the observed panelist value, and μ is the value for the control or target attribute.

ine its ability to perform
uals, as well as the panel
subjects who have other
: sensory program; it is
vation over long periods

alyst needs to have a
el, in terms of validity
e. In certain difference
onal attribute tests, the
reference, the sweeter
r time. The percent of
r monthly or bimonthly
ifficulty of each test in
whole, validity can be
a, instrumental data, or
reatment, addition of a

essed for the individual
g duplicate test samples,

is of variance, the pane-
of the data analysis (see
iled description of this
accepted in QDA that
erceptions of the same
eir absolute values that
Spectrum, panelists are
e scale when evaluating
an analytical instrument
addition to the ASTM
dividual and combined
ance that require moni-
ility). See also Nielsen,

l leader can determine
tributes of a control or
or control rating, called

(9.1)

nd μ is the value for the

Variability. With several evaluations of a blind control or reference, the panelist's variability about his/her own mean rating is calculated using the panelist's standard deviation as follows:

$$\text{panelist SD, } s = \sqrt{\sum_{i-1}^{n} (x_i - \bar{x})^2/(n-1)}. \qquad (9.2)$$

Good panelists have both low bias and low variability. The bias formula may be modified by removing the sign; this produces the absolute bias, calculated as

$$\text{panelist bias, } |d| = |x - \mu|, \qquad (9.3)$$

so that large positive and negative deviations do not offset each other. Small values of absolute bias are desirable. The panelists' statistics should be plotted over time to identify those panelists who need retraining or calibration.

When split-plot analysis of variance is used for descriptive data analysis, the judge-by-sample interaction is part of the results. When this interaction is significant, it is necessary to look at plots of the data to determine the source(s). Figure 9.2 shows three plots of judge-by-sample interactions. In each graph, each line represents one panelist's average ratings for two samples. In the first plot (A), the judge-by-sample interaction is not significant. All judges tend to rate the samples in the same direction and with the same relative degrees of intensity. Thus the lines are in the same direction and similar in slope. The second plot (B) shows an extreme case of judge-by-sample interaction: Several samples are rated quite differently by some of the judges. Consequently, the lines run in different directions and have different slopes. The third plot (C) shows a few judges whose slopes differ from the rest. In this case, although the judge-by-sample interaction is statistically significant, the problem is less extreme. It is one of slight differences in the use of scales rather than total reversals, as in plot B. Generally, a judge-by-sample interaction indicates the need for more training, more frequent use of reference scales, or review of terminology.

9.5.2 Panelist Maintenance, Feedback, Rewards, and Motivation

One of the major sources of motivation for panelists is a sense of doing meaningful work. After a project is completed, panelists should be informed by letter or a posted circular of the project and test objectives, the test results, and the contribution made by the sensory results to the decision taken regarding the product. Immediate feedback after each test also tends to give the individual panelist a sense of "How am I doing?" The fears of some project leaders that panelists might become discouraged in tests with a low probability of success (a triangle test often has fewer than 50% correct responses) have proven groundless. Panelists do take into account the complexity of the sample, the difficulty of the test, and the probability of success. Panelists do want to know about the test, and can indeed learn from past performance. Discussion of results after a descriptive panel session is highly recommended. The need to constantly refine the terms, procedures, and definitions is best served by regular panel interaction after all the data have been collected.

Feedback to panelists on performance can be provided with data regarding their individual performance over three to five repeat evaluations of the same product vis-à-vis the panel as a whole. The data in Table 9.8 for a given sample indicates the mean and standard deviation for each panelist (numbers) for each attribute (letters), as well as the panel mean

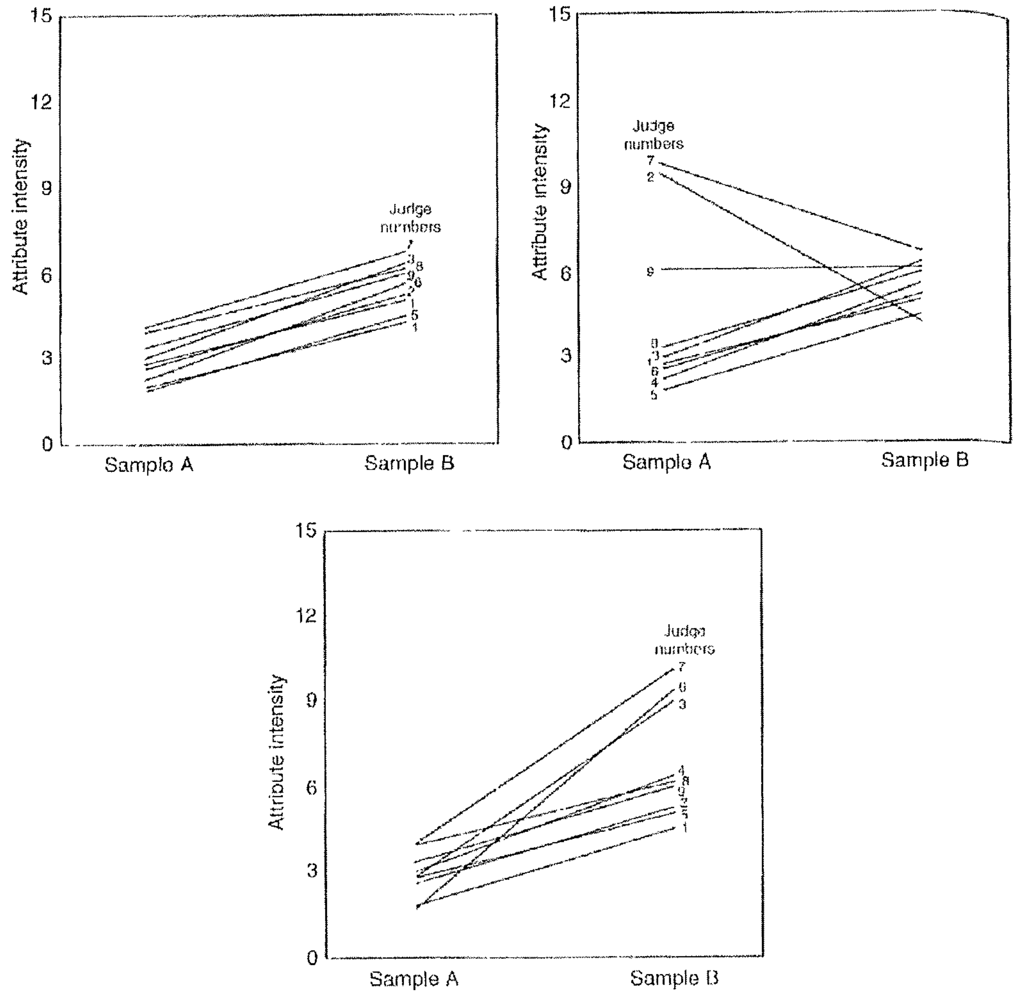


FIGURE 9.2
Judge and sample interaction plots (see text)

and standard deviation. Panelists can then determine how well the individual means agree with that of the panel as a whole (bias). In addition, the panelist's standard deviation provides an indication of that panelist's reliability (variability) on that attribute. Data for two or three products or samples over three to five evaluations should be shown to panelists on a regular basis, e.g., every three to four months. Plots of judge-by-sample interaction, such as those shown in Figure 9.2, may also be shown to panelists to demonstrate both the general agreement among all the panelists and the performance of each panelist relative to the others.

In addition to the psychological rewards derived from feedback, panelists also respond positively and are further motivated to participate enthusiastically by a recognition and/or reward system. The presentation of certificates of achievement for:

- High panel attendance
- High panel performance

TABLE 9.8

Panel Performance Summary

Attributes	Panelist 1	2	3	4	5...	14	Panel X/SD
A	7.5/02[a]	7.0/2	6.8/2	6.9/1	7.9/2.5	6.2/1.9	6.9/05[b]
B	4.2/1 4	4 8/2	5.5/1.6	5.0/0	4.2/1.2	4.6/1.6	4.8/0.4
C	1.4/1	3/1.3	1.5/1.2	1.0/0.9	1.1/0.8	3/1.3	1.8/0.8
D	9.0/0.5	8.0/0.7	9.0/1.0	6.4/1.2	12/1.1	10/1.3	9.4/1.6
E	4.0/0.7	4.2/0.8	3.5/1	1 9/1.2	4.4/0.9	3.8/2	3.9/1.1

The 14 panelists evaluated the same sample in between other samples over a period of 3 weeks. The panel grand mean for attribute A was 6.9 and the SD over the 14 panelist means was 0.5 or 7.2%, showing satisfactory agreement between panelists for this attribute. Panelist 5 rated the attributes A and E much higher than the panel means and showed a high SD for attribute A.

[a] Panelist mean/standard deviation.

[b] Panelist grand mean/grand standard deviation.

- Improved performance
- Completion of a training program
- Completion of a special project

stimulates panel performance and communicates to panelists that the evaluation is recognized as worthwhile. Short-term rewards, such as snacks, tokens for company products, and raffle tickets for larger prizes, are often given to subjects daily. Over the longer term, sensory analysts often sponsor parties, outings, luncheons, or dinners for panelists, if possible, with talks by project or company management describing how the results were used. Publicity for panel work in the company newspaper or the local community media serves to recognize the current panel members and stimulates inquiry from potential candidates. Being a panelist is about discovering all of the sensory nuances the samples display. The ability to discover is strengthened by encouraging the panelists to become more sensory aware. Activities designed to increase sensory awareness are also motivating to the panel. The activities allow the panelists to learn new information while having a bit of fun and further stimulate the mind (Appendix 9.2B). Panel breakdown can occur if the panel leader does not set clear boundaries on acceptable and unacceptable behavior. It is a good idea to establish guidelines for expected behavior with the panel early on. Written guidelines that are reviewed and signed by the panelists serve as the foundation for panel operations (Appendix 9.2C). The underlying support by management for the full sensory program and for the active participation by panelists is a key factor in recruiting and maintaining an active pool of highly qualified members.

Appendix 9.1 Prescreening Questionnaires

Each of the prescreening questionnaires is designed to enable the panel leader or trainer to select from a large group of candidates those individuals who are both